Sketched SVD: Recovering Spectral Features from Compressive Measurements[*]

Anna C. Gilbert
University of Michigan
Ann Arbor, Michigan
annacg@umich.edu

Jae Young Park
University of Michigan
Ann Arbor, Michigan
jaeypark@umich.edu

Michael B. Wakin
Colorado School of Mines
Golden, Colorado
mwakin@mines.edu

ABSTRACT

We consider a streaming data model in which n sensors observe individual streams of data, presented in a turnstile model. Our goal is to analyze the singular value decomposition (SVD) of the matrix of data defined implicitly by the stream of updates. Each column i of the data matrix is given by the stream of updates seen at sensor i. Our approach is to sketch each column of the matrix, forming a "sketch matrix" Y, and then to compute the SVD of the sketch matrix. We show that the singular values and right singular vectors of Y are close to those of X, with small *relative* error. We also believe that this bound is of independent interest in non-streaming and non-distributed data collection settings.

Assuming that the data matrix X is of size $N \times n$, then with m linear measurements of each column of X, we obtain a smaller matrix Y with dimensions $m \times n$. If $m = O(k\epsilon^{-2}(\log(1/\epsilon) + \log(1/\delta)))$, where k denotes the rank of X, then with probability at least $1 - \delta$, the singular values σ'_j of Y satisfy the following relative error result

$$(1 - \epsilon)^{\frac{1}{2}} \leq \frac{\sigma'_j}{\sigma_j} \leq (1 + \epsilon)^{\frac{1}{2}}$$

as compared to the singular values σ_j of the original matrix X. Furthermore, the right singular vectors v'_j of Y satisfy

$$\|v_j - v'_j\|_2 \leq$$
$$\min \left\{ \sqrt{2}, \ \frac{\epsilon\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}} \max_{i \neq j} \frac{\sqrt{2}\sigma_i\sigma_j}{\min_{c \in [-1,1]}\{|\sigma_i^2 - \sigma_j^2(1 + c\epsilon)|\}} \right\},$$

as compared to the right singular vectors v_j of X. We apply this result to obtain a streaming graph algorithm to approximate the eigenvalues and eigenvectors of the graph Laplacian in the case where the graph has low rank (many connected components).

---

[*]Copyright statement goes here.

1. INTRODUCTION

Consider a collection of data arranged in a matrix X of size $N \times n$. Each column represents a length-N signal (or image, frequency counts of terms in a particular document, etc.) and there are n such observations. The singular value decomposition (SVD) of X,

$$X = U\Sigma V^T,$$

carries important information about the structure of the data set, especially when the rank k of X is small. In particular, the columns of U (known as the *left* singular vectors of X) span the principal directions of the data set and can be used as basis vectors for building up typical signals, and the diagonal entries of Σ (known as the singular values of X) reflect the energy of the data set in each of these directions. The extraction of these features is commonly known as Principal Component Analysis (PCA), and PCA is a fundamental and commonly used tool in data analysis and compression. This exact same process can be viewed through a slightly different lens when one imagines the columns of X as independent realizations of a length-N random vector x. Computing the left singular vectors of X is equivalent to computing the eigenvectors of XX^T, which (up to rescaling) is the $N \times N$ sample covariance matrix of the data. In this context, PCA is also known as the Karhunen-Loeve (KL) Transform, and the KL Transform is a fundamental and commonly used tool in statistics.

There are in fact a number of applications where the *right* singular vectors V of a data matrix X are more important, or equivalently, where the eigenvectors of $X^T X$ carry the structure of interest. For example, the product ΣV^T gives a low-dimensional embedding of the data set that preserves distances and angles between the n data vectors. This embedding can be used for clustering or categorizing the signals [6, 27]; for example, this is used for comparing documents in latent semantic analysis. The right singular vectors of X can also be viewed as the result of applying the KL Transform to the rows, rather than the columns, of X. In this sense, the columns of V describe the inter-signal (rather than intra-signal) statistical correlations. In cases where the column index corresponds to a distinct sensor position, or a vertex in a graph, etc., this correlation structure can carry important structural information [8, 31].

Unfortunately, many of the applications in which we seek the right singular vectors of X (equivalently, the eigenvectors of $X^T X$) are those in which the data is simply too large,

too distributed, or generated too quickly for us to store the data or to process it efficiently in one, centralized location. There are, however, settings in which the data sets—while large—have low intrinsic dimension or are of low rank. Let us suppose that the length of each data vector N is much larger than the number of observations n, and suppose that X has rank $k \leq n$. The data may or may not be generated in a dynamic, streaming fashion and it may or may not be collected in a distributed fashion amongst n sensors.

We wish to design a joint observation process (which can be distributed amongst n sensors) that maintains a "sketch" of the data stream and a reconstruction process that, at a central location, reconstructs *not* an approximation of the original data, but rather a good approximation to the singular values σ_j and the (right) singular vectors v_j of the original matrix X. The sketch of the data stream should be a linear, non-adaptive procedure, one that is efficient to update, and one that uses as few observations of the data matrix as possible so that as little communication as possible is required from the sensors to the central processing entity. Because the procedure is linear and non-adaptive, we can represent the sketch as a matrix-matrix product $\Phi X = Y$, where the observation matrix Φ is of size $m \times N$ and the sketch Y is of size $m \times n$. We want m as small as possible. From the sketch matrix Y, we want to produce estimates σ'_j of the k non-zero singular values and estimates v'_j of the associated (right) singular vectors of X such that

$$(1 - \epsilon)^{1/2}\sigma_j \leq \sigma'_j \leq (1 + \epsilon)^{1/2}\sigma_j$$

and

$$\|v_j - v'_j\|_2 \leq \epsilon\gamma,$$

where γ is a function of the smallest gap between the singular values of X.

Most of the current work on low-rank matrix approximations, robust PCA [7], rank-revealing QR factorizations (see [20] and the references therein for a comprehensive survey), etc. has focused on obtaining a good approximation $\widehat{X}$ to the original data matrix X, albeit one that is parsimonious. Some of the measurement schemes [7] sketch both the row and column space of X, collecting measurements $Y = \Phi X \Phi^T$, and some sketch the column space only so as to derive a few orthogonal vectors that span the column space [20]. These results are of the form

$$\|X - \widehat{X}\|_{\mathcal{X}} \leq (1 + \epsilon)\|X\|_{\mathcal{X}}$$

for some norm $\mathcal{X}$ (typically, the Frobenius norm, but others as well). Our goal is an approximation to the *singular vectors* and *singular values* of X themselves, directly, and while one could apply standard perturbation theory techniques to compare the singular vectors of an approximation $\widehat{X}$ to those of the original data X, the error guarantees would be rather poor. Furthermore, while one could ask about preserving the subspaces spanned by the singular vectors, there are many applications from PCA to data clustering, image segmentation, graph embedding, and modal analysis in which the individual singular vectors are critical for data analysis tasks, and data reconstruction is not necessarily required.

Our algorithmic approach is straightforward. We sketch one side of the $N \times n$ data matrix X, maintaining a sketch matrix $Y = \Phi X$ of size $m \times n$. (The fact that the sketch is one-sided allows it to be computed sensor-by-sensor in distributed data collection settings.) We then compute the SVD of the sketch matrix Y, using standard (iterative) SVD algorithms. Our analysis is quite different from that of most randomized linear algebra methods. We assume that the sketching matrix Φ is randomly generated and satisfies the distributional Johnson-Lindenstrauss (JL) property (see Definition 1) so that with high probability it acts as a near isometry on the column span of X, and we then exploit *relative error* (as opposed to absolute error) perturbation analysis for *deterministic* (as opposed to random) matrices to obtain our results. As detailed in our main result, Theorem 1, we can obtain accurate relative estimates for the singular values, and in some cases we can obtain accurate estimates for the singular vectors as well. However, we struggle to achieve high accuracy in the singular vectors when the corresponding singular values are close. This is a consequence of well-studied perturbation theory and seems inherent in our approach.

One major application of our work is to determining structural graph properties of streaming graph data, albeit for low-rank graphs (ones with many connected components). Recent work on the structure and evolution of online social networks [22] suggests that a significant fraction of vertices in such networks participate in isolated communities, "small groups who interact with one another but not with the network at large."

In Section 2, we set the stage for our mathematical problem and in Section 3, we outline the related work, including an overview of the relative error perturbation techniques from linear algebra that we will use. In Section 4, we present our main result which we apply, in Section 5, to the spectral analysis of streaming graphs.

2. PROBLEM SETUP

Let X denote the $N \times n$ real-valued data matrix. For example, one may envision that each column represents a length-N time series signal collected from one of n sensors. Let us assume that $N \geq n$, and that $\text{rank}(X) = k \leq n$. We write the *truncated* SVD of X as $X = U_X \Sigma_X V_X^T$, where unlike the full SVD, U_X is $N \times k$, $\Sigma_X = \text{diag}(\sigma_1, \ldots, \sigma_k)$ with $\sigma_1 \geq \cdots \geq \sigma_k > 0$, and V_X is $n \times k$. Our goal in this paper is to estimate the singular values Σ_X and the right singular vectors V_X from a low-dimensional sketch obtained by left-multiplying X with a compressive matrix. In particular, we let Φ denote a sketching matrix of dimension $m \times N$, and we denote the sketch of X by $Y = \Phi X$. We are specifically interested in cases where $m < N$ and thus Y is a shorter matrix than X. We also note that the sketching matrix Φ can be applied individually to each column of X, and these sketched columns can be concatenated to form the $m \times n$ matrix Y. In other words, the sketching can be performed sensor-by-sensor.

Our algorithm for estimating Σ_X and V_X from Y is explained in Section 4.1. This algorithm is very simple and is based on the idea that under a suitable choice of Φ, the singular values and right singular vectors of Y will approximate the singular values and right singular vectors of X. In order to state our results, we write the truncated SVD of Y as $Y = U_Y \Sigma_Y V_Y^T$, where U_Y is $m \times k$, $\Sigma_Y = $

diag$(\sigma'_1, \ldots, \sigma'_k)$ with $\sigma'_1 \geq \cdots \geq \sigma'_k \geq 0$, and V_Y is $n \times k$. (We will be interested in cases where $m \geq k$, and typically Y will have rank k just like X.) It will also be useful for us to write the eigendecompositions of $X^T X$ and $Y^T Y$ as $X^T X = V_X \Lambda_X V_X^T$ and $Y^T Y = V_Y \Lambda_Y V_Y^T$, respectively, where $\Lambda_X = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ and $\Lambda_Y = \mathrm{diag}(\lambda'_1, \ldots, \lambda'_k)$. (Although it is not common practice we will use the "truncated" eigendecomposition so that V_X and V_Y are both of dimension $n \times k$ and Λ_X and Λ_Y are both of dimension $k \times k$; when $k = n$ we will have the usual eigendecomposition.) Note that $\lambda_j = \sigma_j^2$ and $\lambda'_j = \sigma_j'^2$ for $j = 1, \ldots, k$.

To ensure that the spectral information about X is preserved in the sketched matrix $Y = \Phi X$, we rely on randomness to construct the sketching matrix Φ. Any random distribution that can be used to construct a Johnson-Lindenstrauss (JL) embedding can be used to generate Φ.

DEFINITION 1. *An $m \times N$ random matrix Φ is said to satisfy the* distributional JL property *if for any fixed $x \in \mathbb{R}^N$, and any $0 < \epsilon < 1$,*

$$Pr\left[\left|\|\Phi x\|_2^2 - \|x\|_2^2\right| > \epsilon\|x\|_2^2\right] \leq 2e^{-mf(\epsilon)},$$

where $f(\epsilon) > 0$ is a constant depending only on ϵ.

For most random matrices satisfying the distributional JL property, the functional dependence on ϵ, $f(\epsilon)$, is quadratic in ϵ as $\epsilon \to 0$. For compactness, we suppress this except where necessary for quantifying the number of measurements or the run time of our algorithm. There are a variety of random matrix constructions known to possess the distributional JL property. Notably, random matrices populated with independent and identically distributed (i.i.d.) subgaussian entries will possess this property [12]. Subgaussian distributions include suitably scaled Gaussian and ± 1 Bernoulli random variables. Other examples of non-i.i.d. JL matrices and discussions of the function $f(\epsilon)$ are contained in works such as [11, 21]. There are many papers that address the sparsity of a JL matrix, the speed of its application to a vector, the minimum number of rows it must possess, the minimum amount of randomness necessary to generate such a matrix, etc. For our results, a random matrix satisfying the distributional JL property is sufficient and, depending on the particular application (streaming versus static data), we want either fast update times (i.e., a sparse transform) or a fast transform. We appeal to a long line of work in assessing the qualities of such transforms and in constructing them, either randomly or deterministically.

Finally, we note that a primary objective of this work is to quantify the amount of perturbation of the right singular vectors of X under the random measurement operator Φ. For $j = 1, \ldots, k$, let us denote the jth column of V_X as v_j and the jth column of V_Y as v'_j. Our bounds concern the quantity $\|v_j - v'_j\|_2$. However, we note that the right singular vectors v_j and v'_j are each unique only up to multiplication by -1. Thus, without loss of generality, we will assume that the sign of each v'_j is chosen so that it is positively correlated with v_j. That is, we will assume for each j that $\langle v_j, v'_j \rangle \geq 0$.

3. RELATED WORK

In order to quantify the amount of change in the singular values and the right singular vectors, we approach this problem from the matrix perturbation theoretic perspective. To see the connection to matrix perturbation theory, let us write

$$Y^T Y = X^T \Phi^T \Phi X = V_X \Sigma_X U_X^T \Phi^T \Phi U_X \Sigma_X V_X^T.$$

Defining $\Delta_\Phi := \Phi^T \Phi - I$, $Y^T Y$ becomes

$$Y^T Y = V_X \Sigma_X U_X^T (I + \Delta_\Phi) U_X \Sigma_X V_X^T,$$
$$= V_X \Sigma_X^2 V_X^T + V_X \Sigma_X U_X^T \Delta_\Phi U_X \Sigma_X V_X^T.$$

Given this equation, we can think of the symmetric matrix $Y^T Y$ as being the summation of some original symmetric matrix

$$A := V_X \Sigma_X^2 V_X^T$$

and a perturbation matrix

$$E := V_X \Sigma_X U_X^T \Delta_\Phi U_X \Sigma_X V_X^T.$$

Roughly speaking, when E is small in some sense, it would be reasonable to expect $Y^T Y$ to have approximately the same spectral information as A. Thus we can think about our problem as quantifying the amount of change between the eigenvalues and eigenvectors of A (which equal the squared singular values and the right singular vectors of X) and those of $Y^T Y$ under the additive perturbation E.

In the following sections, we briefly review some of the important results in the matrix perturbation theory literature and also discuss the connection of our problem to the simultaneous iteration method, an important algorithm for computing the eigendecomposition of a matrix.

3.1 Absolute Bounds

There is an extensive literature in the field of matrix perturbation theory quantifying the amount of change in the eigenvalues and eigenvectors of a symmetric $n \times n$ matrix A when it undergoes an additive perturbation. A perturbed matrix, $\widetilde{A}$, may be written in the form of $\widetilde{A} = A + E$, where E is the perturbation matrix that is being added to A. It is well-known that the eigenvalues of A and those of $\widetilde{A}$ will be close to one another when the amount of perturbation E is small (typically with respect to the 2-norm of E). Let us denote the jth largest eigenvalues of A and $\widetilde{A}$ as λ_j and $\widetilde{\lambda}_j$, respectively. Then, it is known [19] that for $j = 1, \ldots, n$,

$$|\widetilde{\lambda}_j - \lambda_j| \leq \|E\|_2. \tag{1}$$

Thus we can see that the distance between each perturbed eigenvalue and the corresponding original eigenvalue will depend on the amount of perturbation, i.e., $\|E\|_2$.

To discuss the perturbation in the eigenvectors let us first represent the eigenvectors of A as v_j such that $Av_j = \lambda_j v_j$. Similarly, let us define the eigenvectors of $\widetilde{A}$ as $\widetilde{v}_j$ such that $\widetilde{A}\widetilde{v}_j = \widetilde{\lambda}_j \widetilde{v}_j$. It is well known that for general matrices A and E, the eigenvectors v_j and $\widetilde{v}_j$ may vary drastically even when the amount of perturbation is small. In other words, $\|\widetilde{v}_j - v_j\|_2$ can be large even when $\|E\|_2$ is small. To see why, let us for example look at the case when two eigenvalues, λ_1 and λ_2, of A are equal to each other. For such a case, we know that the eigenvectors corresponding to those eigenvalues, v_1 and v_2, will not be unique: any linear combination of the two

eigenvectors will also be a valid eigenvector corresponding to the same eigenvalue. A perturbation to this matrix will generally cause λ_1 and λ_2 to split into two eigenvalues $\widetilde{\lambda}_1$ and $\widetilde{\lambda}_2$, each of which will satisfy equation (1) above. If $\widetilde{\lambda}_1$ and $\widetilde{\lambda}_2$ are distinct, the corresponding eigenvectors $\widetilde{v}_1$ and $\widetilde{v}_2$ will now be uniquely identified. Since v_1 and v_2 were not unique, it is possible for $\widetilde{v}_1$ to differ from any particular choice of v_1 and for $\widetilde{v}_2$ to differ from any particular choice of v_2.

The perturbation in the eigenvectors, however, is not completely arbitrary. It is known that the angle between the space spanned by v_1 and v_2 and the space spanned by $\widetilde{v}_1$ and $\widetilde{v}_2$ will be small if E is small. To state this result more concretely, let us represent the eigendecompositions of A and $\widetilde{A}$ as

$$A = V\Lambda V^T = (V_1\ V_2) \begin{pmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_2^T \end{pmatrix}$$

and

$$\widetilde{A} = \widetilde{V}\widetilde{\Lambda}\widetilde{V}^T = \left(\widetilde{V}_1\ \widetilde{V}_2\right) \begin{pmatrix} \widetilde{\Lambda}_1 & 0 \\ 0 & \widetilde{\Lambda}_2 \end{pmatrix} \begin{pmatrix} \widetilde{V}_1^T \\ \widetilde{V}_2^T \end{pmatrix}.$$

The eigenvalue matrices are such that,

$$\Lambda_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_p),\ \Lambda_2 = \mathrm{diag}(\lambda_{p+1}, \ldots, \lambda_n),$$
$$\widetilde{\Lambda}_1 = \mathrm{diag}(\widetilde{\lambda}_1, \ldots, \widetilde{\lambda}_p),\ \widetilde{\Lambda}_2 = \mathrm{diag}(\widetilde{\lambda}_{p+1}, \ldots, \widetilde{\lambda}_n)$$

for an arbitrarily chosen $p \in \{2, \ldots, n-1\}$.

It is possible to quantify how close the spaces spanned by the columns of V_1 and $\widetilde{V}_1$ are. In order to provide a measure of closeness between the two spaces, the following notion of angle matrix was defined in [14]:

$$\Theta(X_1, X_2) :=$$
$$\arccos\left(\left(X_1^T X_1\right)^{-\frac{1}{2}} X_1^T X_2 (X_2^T X_2)^{-1} X_2^T X_1 (X_1^T X_1)^{-\frac{1}{2}}\right)^{-\frac{1}{2}},$$

where X_1 and X_2 are two matrices of the same dimension $n \times p$ with $n > p$ and full column rank. The singular values of $\Theta(X_1, X_2)$ are the angles required to rotate the space spanned by X_1 onto that of X_2. Going back to our notation of V_1, V_2, $\widetilde{V}_1$, and $\widetilde{V}_2$, it was shown [14] that

$$\|\sin\Theta(V_1, \widetilde{V}_1)\| = \|\widetilde{V}_2^T V_1\|$$

for any unitarily invariant norm. This fact can be used to bound the angle between V_1 and $\widetilde{V}_1$. In particular, if $\kappa := \min|\lambda(\Lambda_1) - \lambda(\widetilde{\Lambda}_2)| > 0$, then

$$\|\sin\Theta(V_1, \widetilde{V}_1)\|_F = \|\widetilde{V}_2^T V_1\|_F \leq \frac{\|E\|_F}{\kappa}.$$

Once again we note that the above bound relies on the absolute separation between eigenvalues (in contrast with the relative separation, which appears in Section 3.2). The above can be further generalized to any invariant norm. Detailed discussion on this subject can be found in [14].

3.2 Relative Bounds
The perturbation results discussed above are in terms of the absolute differences between eigenvalues. These types of results are most useful for ensuring the preservation of

the largest eigenvalues but least useful for ensuring the relative preservation of the smallest eigenvalues; a small *absolute* change in a small eigenvalue could actually correspond to a large *relative* change in that eigenvalue. The absolute perturbation results are the best we can do when the perturbation to A is completely arbitrary. However, when the perturbation exhibits some structure one can do much better than what the absolute error bounds indicate.

Consider the class of perturbations that take the form $\widetilde{A} = A + E = D^T A D$, where D is non-singular. It was shown [16] that in this case a relative perturbation bound for the eigenvalues is given by

$$|\widetilde{\lambda}_j - \lambda_j| \leq |\lambda_j|\|D^T D - I\|_2,$$

where the factor $\|D^T D - I\|_2$ represents how close D is to being an orthonormal matrix. In the extreme case when D has orthonormal columns we will have that $\widetilde{\lambda}_j = \lambda_j$ as expected.

Similarly, the angle $0 \leq \theta_j \leq \frac{\pi}{2}$ between the jth eigenvector and its corresponding perturbed eigenvector has been shown [16] to satisfy

$$\sin\theta_j \leq \frac{\|D^T D\|_2 \|(DD^T)^{-1} - I\|_2}{\rho_j(A) - \|D^T D - I\|_2} + \|D - I\|_2,$$

provided that $\rho_j(A) > \|D^T D - I\|_2$, where the jth relative gap, $\rho_j(A)$, of the eigenvalues of A is defined as

$$\rho_j(A) = \min_{i \neq j} \frac{|\lambda_i - \lambda_j|}{|\lambda_j|}.$$

We can see that for the type of perturbation described above we obtain a much stronger perturbation result that depends on the relative gap between the eigenvalues. There are many variants of relative perturbation results that have been proposed to date [5, 16, 23, 24] that differ from one another depending on the underlying matrix A (e.g., whether it is a symmetric matrix, positive definite matrix, indefinite matrix, etc.) and also on the type of perturbation.

3.3 Relation to Simultaneous Iteration
Our problem also has a close connection to various algorithms for computing the eigenvalues and eigenvectors of symmetric matrices. One algorithm that we shall focus on is the simultaneous iteration method [9, 30]. This method is best suited for cases when we are interested in computing the top few eigenvalues and their corresponding eigenvectors and when the underlying matrix is sparse.

To state the algorithm explicitly, let us set some notation. Let A be an $n \times n$ positive-semidefinite matrix with eigendecomposition $A = V\Lambda V^T$. We let λ_j and v_j denote the jth eigenvalue of A and its corresponding eigenvector. We also assume that the eigenvalues are ordered in descending order such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \geq 0$. We then pick a set of trial vectors and denote them as $p_1, p_2, \ldots, p_k$, where the number of trial vectors k depends on how many eigenvectors we wish to compute. The trial vectors can be any set of orthonormal vectors such that

$$\mathrm{span}(p_1, \ldots, p_k) \bigcap \mathrm{span}(v_{k+1}, \ldots, v_n) = \{0\}. \qquad (2)$$

One possible choice of trial vectors is a set of k orthonormal vectors that are chosen randomly. Let us stack the trial vectors into columns of a matrix and denote it as $P = [p_1, \ldots, p_k]$. Given this notation the simultaneous iteration method is carried out as follows:[1]

1. $W^{(0)} \leftarrow AP$

2. for $i = 1, 2, \ldots$

   (a) $Q^{(i)} R^{(i)} \leftarrow W^{(i)}$ via the QR-decomposition

   (b) $W^{(i+1)} \leftarrow AQ^{(i)}$

   (c) if stopping criterion is not met, set $i \leftarrow i+1$ and go back to step (a), otherwise output $Q^{(i)}$.

As we can see, the simultaneous iteration method iteratively refines the set of eigenvectors of A. Denoting the jth column of $Q^{(i)}$ as $q_j^{(i)}$, it is known [30] that

$$\|q_j^{(i)} - v_j\|_2 = \mathcal{O}(p_j^i),$$

where $p_j = \max\{\lambda_{j+1}/\lambda_j, \lambda_j/\lambda_{j-1}\}$. From this we can see that the rate of convergence of the eigenvectors depends on the ratio between the eigenvalues. Put differently, an eigenvector that corresponds to an eigenvalue with a favorable eigenvalue ratio p_j will converge faster to the true eigenvector. This is also similar in spirit to the relative perturbation results that we discussed above in that p_j provides relative measure of the closeness between the eigenvalues, and the accuracy between v_j and $q_j^{(i)}$ depends on p_j.

Lastly, let us focus on the very first step in the simultaneous iteration algorithm, in which we multiply the original matrix A with k (potentially randomly chosen) vectors. Interestingly, this is similar in spirit to our algorithm, except that our data matrix X is not necessarily square or positive-semidefinite, and we do not require the rows of Φ to be orthonormalized. We would like to note that—when they are used in the simultaneous iteration method—random trial vectors are merely chosen as a way to satisfy the condition (2). We believe, however, that randomness will also help to better preserve the true eigenvectors in the first iteration.

3.4 Randomized Algorithms for Linear Algebra

In a similar vein, there have been a large number of results on what we will refer to as randomized algorithms for linear algebra. The monograph [25] covers a number of these methods and references.

There are several lines of work that are closely related to our results. The first involves the spectral analysis of random matrices and the application to algorithmic tasks such as information retrieval and spectral random graph analysis. Two representative works are [3, 10] which build random matrix models (or random perturbations of random matrices) for data and graphs and then use those models to find the approximate spectral structures in the data (or the SVDs). Many of the perturbation results used in these papers fall into our category of absolute bounds.

A second line of work is that of robust PCA or low-rank matrix completion, of which [7] is just one example (there are many other such papers). In this problem, the data matrix X is either sampled or random linear measurements of the form $\Phi X \Phi^T$ are obtained, from which a sparse, low-rank approximation $\widehat{X}$ to the original data matrix is produced. The primary goal of this line of work is to approximate the original data with a parsimonious representation. Our work, in contrast, aims to recover or to approximate the parsimonious representation itself.

A third line of work is the recovery of principal components of a data matrix X from compressive projection measurements [18, 29]. Briefly speaking, in these works, a rectangular data matrix X of size $p \times n$ is considered where $p < n$, i.e., X has more columns than rows. Each column of X represents a data sample, and the objective in these works is to compute the *left* singular vectors of X from compressive projections of the columns of X. However, different random projection matrices are applied to different columns of X. The key differences between the work proposed in [18, 29] and our work are that 1) we are interested in the right singular vectors of X, 2) the data matrix X in our problem is assumed to have more rows than columns, 3) our measurement matrix is a JL matrix, and 4) we apply the same random matrix to every column of X.

Finally, we emphasize the distinction between subspace approximation and the approximation of the singular vectors themselves. Feldman et al. [17] give coreset and sketching algorithms for approximating subspaces spanned by portions of the data set. This problem is also similar to the work of Halko et al. [20], in which one constructs a basis for an approximate subspace spanned by the columns of X from a sketch $Y = \Phi X$ of the data. Finally, we note that the work of Drineas et al. [15] that approximates leverage scores of a matrix is similar in nature to ours but does not produce approximate singular vectors.

4. MAIN RESULT
4.1 Proposed Algorithm and Estimation Bounds

Recall the problem setup discussed in Section 2. Our algorithm for estimating Σ_X and V_X from the sketched matrix $Y = \Phi X$ is stated in Algorithm 1. This algorithm is very simple: we simply return the truncated singular values and right singular vectors of Y.

Algorithm 1: Pseudo-code for sketched SVD.

Input: Sketched matrix $Y = \Phi X$
Output: $\widehat{\Sigma}_X$ and $\widehat{V}_X$ (estimates of the singular values and right singular vectors, respectively, of X)
$(U_Y, \Sigma_Y, V_Y) \leftarrow \mathrm{SVD}(Y)$
$\widehat{\Sigma}_X \leftarrow \Sigma_Y$
$\widehat{V}_X \leftarrow V_Y$

The computational complexity of our algorithm can be divided into two parts. The first part concerns the complexity of computing the sketch Y from Φ and X. If both Φ and X are available at a central processing node, Y can be computed simply by multiplying Φ and X; as discussed in Section 2, there may be a fast algorithm for doing this, depending on the structure of Φ. As we have noted, however, it

---

[1] There are a few variants of the simultaneous iteration method. Here, we use the algorithm presented in [30].

is also possible to compute the sketch column-by-column by applying Φ separately to each column of X; when data is collected in a distributed fashion, this may be the natural way to construct a sketch. Let us denote the computational complexity computing Y as $T_1(m, N, n)$. In distributed scenarios, we will have $T_1(m, N, n) = nT'(m, N)$, where $T'(m, N)$ denotes the computational complexity of one matrix-vector multiplication with Φ.

The second part concerns the complexity of computing the SVD of Y. Using standard techniques, computing the SVD of an $m \times n$ matrix with $m \geq n$ requires $\mathcal{O}(mn^2)$ operations. (When k is very small compared to n, we may have $m < n$ and the SVD of Y can be computed even more efficiently than this.) Combining this fact with the bound on m provided in our main result (see (3)) and assuming $f(\epsilon)$ is quadratic in ϵ, the computational complexity of this second part will be $\mathcal{O}(n^2 k \epsilon^{-2}(\log(1/\epsilon) + \log(1/\delta))$, where δ denotes the failure probability. One can add this cost to $T_1(m, N, n)$ to determine the overall computational complexity, although it is important to stress that the computation of Y may be streaming or distributed over many sensors, while the computation of the SVD of Y may be performed all at once at a central computing node.

We are now ready to state our main result.

THEOREM 1. *Let X be an $N \times n$ matrix with $N \geq n$ and $\mathrm{rank}(X) = k \leq n$, and let $X = U_X \Sigma_X V_X^T$ denote the truncated SVD of X as explained in Section 2. Let $\epsilon \in (0, 1)$ denote a distortion factor and $\delta \in (0, 1)$ denote a failure probability, and suppose Φ is an $m \times N$ random matrix that satisfies the distributional JL property with*

$$m \geq \frac{k \log(42/\epsilon) + \log(2/\delta)}{f(\epsilon/\sqrt{2})}. \tag{3}$$

Let $Y = \Phi X$ denote the sketched matrix, and let $\widehat{\Sigma}_X = \Sigma_Y$ and $\widehat{V}_X = V_Y$ denote the estimated singular vectors and right singular values of X returned by Algorithm 1. Then with probability exceeding $1 - \delta$, $\mathrm{rank}(Y) = k$ and both of the following statements hold:

1. *(Preservation of singular values) For all $j = 1, \ldots, k$,*

$$(1 - \epsilon)^{1/2} \leq \frac{\sigma_j'}{\sigma_j} \leq (1 + \epsilon)^{1/2},$$

   where $\sigma_1 \geq \cdots \geq \sigma_k \geq 0$ denote the singular values of X (the diagonal entries of Σ_X) and $\sigma_1' \geq \cdots \geq \sigma_k' \geq 0$ denote the singular values of Y (the diagonal entries of $\widehat{\Sigma}_X$).

2. *(Preservation of right singular vectors) For all $j = 1, \ldots, k$,*

   $\|v_j - v_j'\|_2 \leq$

$$\min \left\{ \sqrt{2}, \ \frac{\epsilon\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}} \max_{i \neq j} \frac{\sqrt{2}\sigma_i\sigma_j}{\min\limits_{c \in [-1,1]}\{|\sigma_i^2 - \sigma_j^2(1+c\epsilon)|\}} \right\},$$

   where $v_1, \ldots, v_k$ denote the right singular vectors of X (the columns of V_X) and $v_1', \ldots, v_k'$ denote the right singular vectors of Y (the columns of $\widehat{V}_X$).

Proof: See Section 4.2.

COROLLARY 1. *When Φ is generated randomly from an i.i.d. subgaussian distribution (suitably scaled) or some other random distribution satisfying the distributional JL property with quadratic $f(\cdot)$, the bounds in Theorem 1 will hold with $m = \mathcal{O}(k\epsilon^{-2}(\log(1/\epsilon) + \log(1/\delta))$.*

This result states that from Y we can obtain accurate relative estimates for the singular values of X, and in some cases we can obtain accurate estimates for the right singular vectors of X as well. However, we struggle to achieve high accuracy in the singular vectors when the corresponding singular values are close. This is a consequence of well-studied perturbation theory (recall the role that the relative gap played in Section 3.2) and seems inherent in our approach.

We note that, naturally, we could obtain similar results for preserving the left singular vectors of X were we to sketch its rows, rather than its columns. We also note that when the exact rank k of the data matrix is unknown (or if the rank of X is not necessarily below n), substituting n for k in the measurement bound (3) yields a guarantee that applies to any $N \times n$ matrix X with $N \geq n$.

4.2 Proof of Theorem 1
In this section we prove Statements 1 and 2 within Theorem 1. As noted in Section 2, it will be useful for us to write the truncated eigendecompositions of $X^T X$ and $Y^T Y$ as $X^T X = V_X \Lambda_X V_X^T$ and $Y^T Y = V_Y \Lambda_Y V_Y^T$, respectively, where $\Lambda_X = \mathrm{diag}(\lambda_1, \ldots, \lambda_k)$ and $\Lambda_Y = \mathrm{diag}(\lambda_1', \ldots, \lambda_k')$. Recall that $\lambda_j = \sigma_j^2$ and $\lambda_j' = \sigma_j'^2$ for $j = 1, \ldots, k$.

4.2.1 Proof of Statement 1
In order to prove Theorem 1, we require the following result, adapted from Lemma 5.1 in [4] and Theorem 4.3 in [13].

LEMMA 1 ([4, 13]). *Let $\mathcal{X}$ denote a k-dimensional subspace of $\mathbb{R}^N$. Let $\epsilon \in (0, 1)$ denote a distortion factor and $\delta \in (0, 1)$ denote a failure probability, and suppose Φ is an $m \times N$ random matrix that satisfies the distributional JL property with*

$$m \geq \frac{k \log(42/\epsilon) + \log(2/\delta)}{f(\epsilon/\sqrt{2})}.$$

Then with probability exceeding $1 - \delta$,

$$\sqrt{1-\epsilon}\|x\|_2 \leq \|\Phi x\|_2 \leq \sqrt{1+\epsilon}\|x\|_2,$$

for all $x \in \mathcal{X}$.

To see how this lemma can help us guarantee the preservation of the singular values of X (or, equivalently, the eigenvalues of $X^T X$), we begin by noting that

$$Y^T Y = X^T \Phi^T \Phi X = V_X \Sigma_X U_X^T \Phi^T \Phi U_X \Sigma_X V_X^T.$$

We define a new $k \times k$ matrix

$$M := V_X^T Y^T Y V_X = \Sigma_X U_X^T \Phi^T \Phi U_X \Sigma_X$$

and represent its eigendecompostion as $M = V_M \Lambda_M V_M^T$. Noting that $Y^T Y = V_Y \Sigma_Y^2 V_Y^T$, we have

$$M = V_X^T Y^T Y V_X = V_X^T V_Y \Sigma_Y^2 V_Y^T V_X.$$

From this we can infer that $\Lambda_M = \Sigma_Y^2$, i.e., M has the same set of eigenvalues as $Y^T Y$. Thus, we turn our attention to proving that the eigenvalues of M approximate the eigenvalues of $X^T X$. Let us define $\Delta_\Phi := \Phi^T \Phi - I$ and consider the ratio

$$\begin{aligned}
\frac{x^T M x}{x^T \Sigma_X^2 x} &= \frac{x^T \Sigma_X U_X^T \Phi^T \Phi U_X \Sigma_X x}{x^T \Sigma_X^2 x} \\
&= \frac{x^T \Sigma_X (I + U_X^T \Delta_\Phi U_X) \Sigma_X x}{x^T \Sigma_X^2 x} \\
&= 1 + \frac{x^T \Sigma_X U_X^T \Delta_\Phi U_X \Sigma_X x}{x^T \Sigma_X^2 x}.
\end{aligned}$$

We will be interested in the range of values that the fraction $(x^T M x)/(x^T \Sigma_X^2 x)$ can take over all nonzero $x \in \mathbb{R}^k$. We note that for any vector $x \in \mathbb{R}^k$ we can associate a vector $y := \Sigma_X x \in \mathbb{R}^k$ and write

$$\frac{x^T M x}{x^T \Sigma_X^2 x} = 1 + \frac{y^T U_X^T \Delta_\Phi U_X y}{y^T y}.$$

To bound the range of values that this quantity can take, it suffices to consider all vectors $y \in \mathbb{R}^k$ with unit norm. Thus, we focus on the quantity

$$\begin{aligned}
1 + y^T U_X^T \Delta_\Phi U_X y &= 1 + y^T U_X^T (\Phi^T \Phi - I) U_X y \\
&= y^T U_X^T \Phi^T \Phi U_X y = \|\Phi U_X y\|_2^2. \quad (4)
\end{aligned}$$

Our next step is to apply Lemma 1 on the subspace $\mathcal{X} = \text{colspan}(U_X)$, using the fact that $\|U_X y\|_2 = \|y\|_2 = 1$. This tells us that with a probability of at least $1 - \delta$,

$$1 - \epsilon \le \|\Phi U_X y\|_2^2 \le 1 + \epsilon,$$

holds for all unit norm vectors $y \in \mathbb{R}^k$. Combining this inequality with (4) we get

$$-\epsilon \le y^T U_X^T \Delta_\Phi U_X y \le \epsilon, \quad (5)$$

which implies that for any nonzero $x \in \mathbb{R}^k$,

$$1 - \epsilon \le \frac{x^T M x}{x^T \Sigma_X^2 x} \le 1 + \epsilon. \quad (6)$$

In order to complete the proof we use the following lemma, which is a simplification of Lemma 1 in [5].

LEMMA 2 (LEMMA 1, [5]). *Let H be a diagonal matrix and suppose δH has the property that for all nonzero x,*

$$g_l \le \frac{x^T (H + \delta H) x}{x^T H x} \le g_u,$$

where $0 < g_l \le g_u$. Then

$$g_l \le \frac{\lambda_i (H + \delta H)}{\lambda_i (H)} \le g_u$$

for all i, where $\lambda_i(Z)$ denotes the ith eigenvalue of the matrix Z.

Applying Lemma 2 to (6) with $H = \Sigma_X^2$ (which is diagonal) and $H + \delta H = M$ completes the proof of Statement 1 of Theorem 1 and also implies that $\text{rank}(Y) = k$.

4.2.2 Proof of Statement 2

In order to prove Statement 2 of Theorem 1, we require the following important theorem.

THEOREM 2 (THEOREM 1, [26]). *Let $H = U \Gamma U^*$ and $\widetilde{H} = H + \delta H = \widetilde{U} \widetilde{\Gamma} \widetilde{U}^*$ be $p \times p$ positive definite matrices. Assume that U and $\widetilde{U}$ are unitary and that $\Gamma = diag(\gamma_1, \dots, \gamma_p)$ and $\widetilde{\Gamma} = diag(\widetilde{\gamma}_1, \dots, \widetilde{\gamma}_p)$ are diagonal. Let $S = U^* \widetilde{U}$, and assume*

$$\eta = \|H^{-\frac{1}{2}} \delta H H^{-\frac{1}{2}}\| < 1,$$

where $H^{-\frac{1}{2}} = U \Gamma^{-1/2} U^*$. Then for any j and for any set $\mathscr{T}$ not containing j we have,

$$\left(\sum_{i \in \mathscr{T}} |s_{ij}|^2 \right)^{1/2} \le \min \left\{ 1, \ \max_{i \in \mathscr{T}} \frac{\gamma_i^{1/2} \widetilde{\gamma}_j^{1/2}}{|\gamma_i - \widetilde{\gamma}_j|} \frac{\eta}{\sqrt{1 - \eta}} \right\},$$

and, in particular, for any $i \ne j$,

$$|s_{ij}| \le \min \left\{ 1, \ \frac{\gamma^{1/2} \widetilde{\gamma}_j^{1/2}}{|\gamma_i - \widetilde{\gamma}_j|} \frac{\eta}{\sqrt{1 - \eta}} \right\}.$$

To prove Statement 2, we continue from the proof of Statement 1. In particular, we suppose (5) holds for all unit norm vectors $y \in \mathbb{R}^k$ (recall that this event happens with probability at least $1 - \delta$). Our first goal will be to prove that this implies that

$$\|U_X^T \Delta_\Phi U_X\|_2 \le \epsilon. \quad (7)$$

To see why this follows, let us for notational simplicity denote $A = U_X^T \Delta_\Phi U_X$. Note that A is a symmetric matrix. Then, (5) says that $-\epsilon \le y^T A y \le \epsilon$ holds for all unit norm vectors $y \in \mathbb{R}^k$. Note that this is equivalent to $-\epsilon \le \frac{y^T A y}{y^T y} \le \epsilon$. This fraction is the well known Rayleigh quotient. It can be shown that the range of values that the Rayleigh quotient takes is confined between the minimum and the maximum eigenvalues of A. Let us denote the maximum and minimum eigenvalues of A as $\alpha_{\max}$ and $\alpha_{\min}$, respectively. Since equation (5) says that the Rayleigh quotient is in between $-\epsilon$ and ϵ we can infer that $-\epsilon \le \alpha_{\min} \le \frac{y^T A y}{y^T y} \le \alpha_{\max} \le \epsilon$. Thus, $\|A\|_2 = \max\{|\alpha_{\min}|, |\alpha_{\max}|\} \le \epsilon$ and so we have proved that (7) holds.

To quantify $\|v_j - v_j'\|_2$, we look at a different yet equivalent quantity that may simplify the problem. Let us again look at the matrix M that we introduced in the proof of Statement 1. We have seen that there is a close connection between the eigenvalues of M and those of $Y^T Y$. We now show that in order to prove that the eigenvectors of $Y^T Y$ approximate those of $X^T X$, it suffices to study the eigenvectors of M.

Remembering that

$$M = V_X^T Y^T Y V_X = V_X^T V_Y \Sigma_Y^2 V_Y^T V_X,$$

we can see that the eigenvectors of M are closely related to the right singular vectors of X and Y. Specifically, we have that $V_M = V_X^T V_Y$, and denoting the jth eigenvector of M as $\widetilde{v}_j$, it is easy to see that $\widetilde{v}_j = V_X^T v_j'$. This implies that $\langle \widetilde{v}_j, e_j \rangle = \langle v_j, v_j' \rangle$ for $j = 1, \dots, k$, where e_j represents the jth canonical basis vector. Furthermore, we note that

colspan(V_Y) = rowspan(Y) = rowspan(X) since every row in Y is a linear combination of the rows in X and since we have argued above that rank(Y) = rank(X). From this (and the fact that $v_j' \in$ colspan(V_Y)) it follows that $\|\widetilde{v}_j\|_2 = \|V_X^T v_j'\|_2 = 1$. Now, using the relation $\langle \widetilde{v}_j, e_j \rangle = \langle v_j, v_j' \rangle$ and the facts that $\|\widetilde{v}_j\|_2 = \|e_j\|_2 = \|v_j\|_2 = \|v_j'\|_2 = 1$, we see that $\|\widetilde{v}_j - e_j\|_2 = \|v_j - v_j'\|_2$. To make sense out of the quantity $\|\widetilde{v}_j - e_j\|_2$, let us examine the expression $M = \Sigma_X^2 + \Sigma_X U_X^T \Delta_\Phi U_X \Sigma_X$. We can view M as the sum of a diagonal matrix Σ_X^2 and a perturbation matrix $\Sigma_X U_X^T \Delta_\Phi U_X \Sigma_X$. The eigenvectors of Σ_X^2 are the canonical basis vectors e_j. Therefore, the quantity $\|\widetilde{v}_j - e_j\|_2$ reflects the amount of perturbation in the eigenvectors of M. This is why, to bound $\|v_j - v_j'\|_2$, it suffices to focus on the perturbation analysis of M.

We apply Theorem 2 as follows: As we have discussed, we will quantify $\|v_j - v_j'\|_2$ via $\|\widetilde{v}_j - e_j\|_2$. Let us set the original unperturbed matrix as $H = \Sigma_X^2$ and the perturbation to this matrix as $\delta H = \Sigma_X U_X^T \Delta_\Phi U_X \Sigma_X$, such that

$$\widetilde{H} = H + \delta H = \Sigma_X(I + U_X^T \Delta_\Phi U_X)\Sigma_X = M,$$

and both H and $\widetilde{H}$ are $k \times k$. Clearly, H is positive definite since it is a diagonal matrix with all positive entries along the diagonal (because rank$(X) = k$). To check that M is positive definite, note that $M = \Sigma_X U_X^T \Phi^T \Phi U_X \Sigma_X$ is of the form $M = B^T B$, where $B = \Phi U_X \Sigma_X$ is an $m \times k$ matrix with $m \geq k$. The fact that B has full column rank will follow because all diagonal entries of Σ_X are nonzero (again, because rank$(X) = k$) and because in the proof of Statement 1 we applied Lemma 1 on the subspace $\mathcal{X} = $ colspan(U_X). Because B has full column rank, M will be positive definite.

We further have that

$$\eta = \|H^{-\frac{1}{2}}\delta H H^{-\frac{1}{2}}\|_2 = \|\Sigma_X^{-1}\delta H \Sigma_X^{-1}\|_2 = \|U_X^T \Delta_\Phi U_X\|_2.$$

Then, applying (7), $\eta = \|U_X^T \Delta_\Phi U_X\|_2 \leq \epsilon$. Let us set $S = I^T V_M = V_M$ and denote the jth eigenvalue of M as $\widetilde{\lambda}_j$. Then, straightforward application of Theorem 2 yields

$$\left(\sum_{i\neq j}|s_{ij}|^2\right)^{1/2} \leq \min\left\{1, \ \max_{i\neq j}\frac{\sigma_i \widetilde{\lambda}_j^{1/2}}{|\sigma_i^2 - \widetilde{\lambda}_j|}\frac{\eta}{\sqrt{1-\eta}}\right\},$$

$$\leq \min\left\{1, \ \max_{i\neq j}\frac{\sigma_i \widetilde{\lambda}_j^{1/2}}{|\sigma_i^2 - \widetilde{\lambda}_j|}\frac{\epsilon}{\sqrt{1-\epsilon}}\right\}.$$

As we have discussed in Section 2 we assume that $s_{jj} = \langle \widetilde{v}_j, e_j \rangle = \langle v_j, v_j' \rangle \geq 0$. Then,

$$\|v_j - v_j'\|_2 = \|\widetilde{v}_j - e_j\|_2 = \sqrt{\|\widetilde{v}_j\|_2^2 - 2\langle \widetilde{v}_j, e_j \rangle + \|e_j\|_2^2},$$

$$= \sqrt{2}\sqrt{1 - s_{jj}},$$

$$= \sqrt{2}\sqrt{1 - \sqrt{1 - \sum_{i\neq j}|s_{ij}|^2}},$$

$$\leq \sqrt{2}\sqrt{1 - \sqrt{1 - \min\left\{1, \ \max_{i\neq j}\frac{\sigma_i^2 \widetilde{\lambda}_j}{(\sigma_i^2 - \widetilde{\lambda}_j)^2}\frac{\epsilon^2}{1-\epsilon}\right\}}},$$

$$\leq \sqrt{2}\min\left\{1, \ \max_{i\neq j}\frac{\sigma_i \widetilde{\lambda}_j^{1/2}}{|\sigma_i^2 - \widetilde{\lambda}_j|}\frac{\epsilon}{\sqrt{1-\epsilon}}\right\},$$

where the last inequality is due to the fact that $1 - \sqrt{1-x} \leq x$ for $0 \leq x \leq 1$.

To write the above only in terms of the unperturbed singular values, σ_j, we make use of Statement 1 and the fact that $\widetilde{\lambda}_j = (\sigma_j')^2$ for $j = 1, \ldots, k$ to obtain

$$\|v_j - v_j'\|_2 \leq$$

$$\min\left\{\sqrt{2}, \frac{\epsilon\sqrt{1+\epsilon}}{\sqrt{1-\epsilon}}\max_{i\neq j}\frac{\sqrt{2}\sigma_i\sigma_j}{\min_{c\in[-1,1]}\{|\sigma_i^2 - \sigma_j^2(1+c\epsilon)|\}}\right\}.$$

5. APPLICATION TO SPECTRAL ANALYSIS OF STREAMING GRAPHS

In this section, we apply our data analysis framework to streaming graphs, a model of data collection where edges of a graph are updated dynamically. We consider a scenario in which edges are inserted and deleted over an observation period, and our goal is to maintain a small data structure that encodes the graph information so that we may analyze the spectrum of the graph quickly at any point during or after the sequence of edge updates. Spectral graph analysis has a multitude of applications including graph embedding, graph isomorphism testing, data clustering/segmentation (of which there are yet many more applications!), numerical linear algebra, etc. We refer the reader to just a few in [6, 8, 27, 31]. Determining the spectrum of a graph is at the heart of many modern data analysis and graphical information processing algorithms.

Let $G = (V, E)$ be a graph with vertex set V and undirected, unweighted edges E. Let A denote the symmetric binary adjacency matrix of G, denote by d_v the degree of a vertex $v \in V$, and define the graph Laplacian as

$$L_G(u,v) = \begin{cases} d_v & \text{if } u = v \\ -1 & \text{if } u \text{ and } v \text{ are adjacent} \\ 0 & \text{otherwise.} \end{cases}$$

There is a compact definition of L_G using the adjacency matrix: $L_G = \text{diag}(d_v) - A$.

Let X be the incidence matrix of the graph G. This matrix has $N = |E|$ rows and $n = |V|$ columns and to define each entry of X, consider an edge (u,v) between vertices u and v. Since the graph is undirected, the ordering of the vertices is chosen arbitrarily. Then,

$$X_{(u,v),u} = 1 \quad \text{and} \quad X_{(u,v),v} = -1.$$

It is well-known that the rank of the graph is the rank of the incidence matrix and that this value is $|V| - c$ where c is the number of connected components in G. If the graph G is weighted, we replace the ± 1's with the appropriate weights in the incidence matrix.

From the definitions of the graph Laplacian and the incidence matrix, it is clear that $L_G = X^T X$. The singular values of X are, therefore, related to the eigenvalues of L_G in a straightforward fashion:

$$\sigma_i(X) = \sqrt{\lambda_i(L_G)}.$$

Furthermore, the *right* singular vectors V of $X = U_X \Sigma_X V_X^T$ are the eigenvectors of the Laplacian. Thus, it is sufficient

to compute (good) approximations to the singular values and the right singular vectors of X to obtain (good) approximations to the top eigenvalues of L_G and the corresponding eigenvectors. From standard spectral graph theory, we know that the eigenvalues λ_i of the Laplacian satisfy $\lambda_1 \geq \cdots \geq \lambda_n = 0$ and, with the assumption that G has c connected components,

$$\lambda_1 \geq \cdots \geq \lambda_{n-c} > \lambda_{n-c+1} = \cdots = \lambda_n = 0.$$

In particular, $\mathrm{rank}(X) = \mathrm{rank}(L_G) = n - c$.

Next, we define the streaming graph model. Following [1, 2], we define a dynamic graph stream as a stream of edge updates (both insertions and deletions). This is a faithful model of the evolution of an online social network, for example, in which users connect and disconnect to other users over time [22].

DEFINITION 2 (DYNAMIC GRAPH STREAM). *A stream $S = \langle a_1, \ldots, a_T \rangle$ where $a_t = (j_t, k_t, \Delta_t) \in [n] \times [n] \times \mathbb{R}$ defines a weighted graph $G = (V, E)$ where $V = [n]$ and the weight of an edge (j, k) is given by*

$$A(j, k) = \sum_{t : (j_t, k_t) = (j,k) \ or \ (k,j)} \Delta_t.$$

We assume that at any update time t, the adjacency matrix A is well-formed; that the edge weight is non-negative; and that the graph has no self-loops.

In this prototype application, the stream of edge updates S defines the edge-vertex incidence matrix X of the graph G. The matrix X has $N = \binom{n}{2}$ rows and n columns, and for each stream item, we update two entries in X as

$$X_{(j_t, k_t), j_t} = X_{(j_t, k_t), j_t} + \Delta_t,$$
$$X_{(j_t, k_t), k_t} = X_{(j_t, k_t), k_t} - \Delta_t.$$

We collect sketches of each column of X and aggregate them into a matrix Y. Denoting the jth column of X by x_j, the jth column of the sketched matrix is given by $y_j = \Phi x_j$. We can update the sketch in a streaming fashion. Upon receipt of a stream item (u_t, v_t, Δ_t), we update y_{u_t} and y_{v_t}:

$$y_{u_t} = y_{u_t} + \Delta_t \phi_{(u_t, v_t)},$$
$$y_{v_t} = y_{v_t} - \Delta_t \phi_{(u_t, v_t)},$$

where ϕ_j denotes the jth column of Φ.

Our main result, Theorem 1, tells us that a sketch of the matrix X is sufficient to recover information about its singular value decomposition.

COROLLARY 2. *Assume that the undirected, weighted graph $G = (V, E)$ is presented in a streaming fashion so that its incidence matrix X has $n = |V|$ columns, $N = \binom{n}{2}$ rows, and rank $k \leq n - 1$. Let $\epsilon \in (0, 1)$ denote a distortion factor and $\delta \in (0, 1)$ denote a failure probability, and suppose Φ is an $m \times N$ random matrix that satisfies the distributional JL property with*

$$m \geq \frac{k \log(42/\epsilon) + \log(2/\delta)}{f(\epsilon/\sqrt{2})}.$$

Let $Y = \Phi X$ denote an $m \times n$ sketch of X maintained in the streaming graph model, and let $\widehat{\Sigma}_X = \Sigma_Y$ and $\widehat{V}_X = V_Y$ denote the estimated singular vectors and right singular values of X returned by Algorithm 1. Then with probability at least $1 - \delta$, the following statements hold:

1. *(Preservation of eigenvalues) For all $j = 1, \ldots, k$,*

$$1 - \epsilon \leq \frac{\lambda_j'}{\lambda_j} \leq 1 + \epsilon$$

   where λ_j denote the true eigenvectors of the graph Laplacian L_G and λ_j' denote the estimated eigenvalues obtained by squaring the diagonal entries of $\widehat{\Sigma}_X$.

2. *(Preservation of eigenvectors) For all $j = 1, \ldots, k$,*
   $\|v_j - v_j'\|_2 \leq$

$$\min \left\{ \sqrt{2}, \ \frac{\epsilon \sqrt{1 + \epsilon}}{\sqrt{1 - \epsilon}} \max_{i \neq j} \frac{\sqrt{2} \lambda_i^{1/2} \lambda_j^{1/2}}{\min_{c \in [-1,1]} \{ |\lambda_i - \lambda_j (1 + c\epsilon)| \}} \right\},$$

   where v_j are the eigenvectors of the graph Laplacian L_G and v_j' denote the estimated eigenvectors obtained from the columns of $\widehat{V}_X$.

Because the adjacency matrix A of G has at most $|V|^2$ nonzero entries, this result is useful only when the rank k of L_G is significantly smaller than $n = |V|$, the number of vertices; or, equivalently, when G has many connected components. In this case, the size of the sketch is smaller than that of the adjacency matrix. In summary, for highly disconnected graphs presented in a streaming fashion, we can recover the approximate eigenvalues and eigenvectors of the Laplacian. The sparsity of the matrix Φ and the speed with which we can update the sketch matrix Y under a stream of updates are functions of the quality of the JL transform. The structural evolution of online social networks [22] suggests that it is reasonable to assume that the underlying graph has a significant fraction of vertices in small, disconnected components so that the graph is essentially a low-rank graph.

6. CONCLUSION

We present a data collection and analysis scheme that permits the distributed collection of data X by resource constrained sensors in a network and the central computation of the spectral decomposition of $X^T X$ or the right singular vectors of the data itself. The algorithm returns not an approximation to the original data, but a good approximation to the singular values σ_j and the right singular vectors v_j of the data. This data collection and analysis framework makes a small number of linear, non-adaptive measurements of the data. The number of measurements each sensor makes is comparable to the rank of the data and, if the data are full rank, the number of measurements at each sensor is comparable to the total number of sensors. This efficient data collection is especially important for sensors that are severely resource constrained and cannot store or transmit a large amount of data to a central device; we believe that one possible application of such an algorithm would be in operational modal analysis of structures (buildings, bridges, etc.) [28].

8. REFERENCES

[1] K. J. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *SODA '12: Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, Jan. 2012.

[2] K. J. Ahn, S. Guha, and A. McGregor. Graph sketches: sparsification, spanners, and subgraphs. In *PODS '12: Proceedings of the 31st symposium on Principles of Database Systems*, May 2012.

[3] Y. Azar, A. Fiat, A. Karlin, F. McSherry, and J. Saia. Spectral analysis of data. In *STOC '01: Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*, 2001.

[4] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constructive Approximation*, 28(3):253–263, 2008.

[5] J. Barlow and J. Demmel. Computing accurate eigensystems of scaled diagonally dominant matrices. Technical report, 1980.

[6] M. Belkin and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems*, 14:585–591, 2001.

[7] E. Candes, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM*, 58(3), 2011.

[8] F. Chung. *Spectral graph theory*, volume 92. Amer Mathematical Society, 1997.

[9] M. Clint and A. Jennings. The evaluation of eigenvalues and eigenvectors of real symmetric matrices by simultaneous iteration. *Comput. J.*, 13(1):76–80, 1970.

[10] A. Dasgupta, J. E. Hopcroft, and F. McSherry. Spectral analysis of random graphs with skewed degree distributions. In *FOCS '04: Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, 2004.

[11] A. Dasgupta, R. Kumar, and T. Sarlos. A sparse Johnson-Lindenstrauss transform. In *STOC '10: Proceedings of the 42nd ACM Symposium on Theory of Computing*, 2010.

[12] M. Davenport. Concentration of measure and sub-gaussian distributions, 2009. Available online at http://cnx.org/content/m32583/latest/.

[13] M. Davenport. *Random Observations on Random Observations: Sparse Signal Acquisition and Processing*. PhD thesis, Rice University, 2010.

[14] C. Davis and W. M. Kahan. The rotation of eigenvectors by a perturbation, III. *SIAM J. Numer. Anal.*, 7, Mar. 1970.

[15] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff. Fast approximation of matrix coherence and statistical leverage. *CoRR*, abs/1109.3843, 2011.

[16] S. C. Eisenstat and I. C. F. Ipsen. Relative perturbation techniques for singular value problems. *SIAM J. Numer. Anal.*, 32(6):1972–1988, Dec. 1995.

[17] D. Feldman, M. Monemizadeh, C. Sohler, and D. P. Woodruff. Coresets and sketches for high dimensional subspace approximation problems. In *SODA '10: Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*, 2010.

[18] J. E. Fowler. Compressive-projection principal component analysis. *IEEE Transactions on Image Processing*, 18(10):2230–2242, 2009.

[19] G. H. Golub and C. F. van Van Loan. *Matrix Computations (Johns Hopkins Studies in Mathematical Sciences)(3rd Edition)*. The Johns Hopkins University Press, 3rd edition, Oct. 1996.

[20] N. Halko, P. Martinsson, and J. Tropp. Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Review*, 53(2):217–288, 2011.

[21] T. S. Jayram and D. Woodruff. Optimal bounds for Johnson-Lindenstrauss transforms and streaming problems with sub-constant error. In *SODA '11: Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms*, 2011.

[22] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In P. S. S. Yu, J. Han, and C. Faloutsos, editors, *Link Mining: Models, Algorithms, and Applications*, pages 337–357. Springer New York, 2010.

[23] R. Li. Relative perturbation theory: (II) Eigenspace and singular subspace variations. *SIAM J. Matrix Anal. Appl*, 20:471–492, 1996.

[24] R. Li. Relative perturbation theory: (I) Eigenvalue and singular value variations. *SIAM J. Matrix Anal. Appl*, 19:956–982, 1998.

[25] M. W. Mahoney. Randomized algorithms for matrices and data. *Found. Trends Mach. Learn.*, 3(2):123–224, Feb. 2011.

[26] R. Mathias and K. Veselić. A relative perturbation bound for positive definite matrices. *Linear Algebra and its Applications*, 270:315 – 321, 1998.

[27] A. Ng, M. Jordan, Y. Weiss, et al. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 2:849–856, 2002.

[28] B. Peeters and G. D. Roeck. Stochastic system identification for operational modal analysis: A review. *Journal of Dynamic Systems, Measurement, and Control*, 123(4):659–667, 2001.

[29] H. Qi and S. Hughes. Invariance of principal components under low-dimensional random projection of the data. *IEEE International Conference on Image Processing (ICIP)*, October 2012.

[30] H. Rutishauser. Computational aspects of F. L. Bauer's simultaneous iteration method. *Numerische Mathematik*, 13:4–13, 1969. 10.1007/BF02165269.

[31] D. Spielman. Spectral graph theory. *Lecture Notes, Yale University*, 2009.